THE CHANGING REQUIREMENTS OF AN ARCHAEOLOGICAL DATA BASE

B.K.W. Booth

Archaeological Research Centre, National Maritime Museum, Greenwich, London SE10

Introduction

This paper examines the varied tasks which an archaeological data base is required to perform; from the initial collection of information during fieldwork, to publication and archiving. In particular data collection in the field, analysis during and after fieldwork, and requirements for output are reviewed. Whilst there are now several working systems for archaeological data (Benson and Jefferies 1980, Flude et al 1981, Catton et al 1981, Griffiths, this volume) there seems to have been little discussion of overall requirements. It is hoped that this paper will generate some discussion on the problems of archaeological data, and how they can be solved. As we are now well on the way to becoming 'automated archaeologists' (Wilcock 1978) it would seem to be sensible to define our aims, priorities and weaknesses, so as to make the most of the exciting new facilities which are now available.

The authors viewpoint is unashamedly that of a micro-computer user. However whilst it is becoming increasingly true that the majority of archaeological computing is on small machines, the paper by Griffiths (this volume) has shown that extremely effective work can be done on mainframe computers. The term 'fieldwork' is used as an alternative to 'excavation', as there are many forms of survey which generate large quantities of data (Pryor 1980). The availability of computer techniques has meant that much of the ordering of data which has traditionally been done after fieldwork may now be performed during fieldwork. Thus the term post-excavation ceases to be relevant, and analysis during and after fieldwork are here considered together.

As an example of an approach to organising archaeological data current work in the Archaeological Research Centre at the National Maritime Museum is cited. The overall scheme for this project, and its predecessor at Maxey, are discussed in two forthcoming papers (Booth, Booth et al).

Data collection during fieldwork

Data collection has progressed to the now ubiquitous printed form, replacing the traditional scruffy notebook with a medium which is more structured, but just as

vulnerable to the elements. If computer methods are to be used the usual procedure is for data to be gathered on forms, and entered into the computer afterwards. There are two disadvantages to this routine. Firstly the transfer of data from the printed form to the computer introduces the possibility of error, particularly if the person inputting the data is not familiar with the material. Secondly conditions may make it difficult to fill out forms; for instance in rain or high winds, or underwater. Waterproof paper is one solution to this problem.

The idea of using a computer during fieldwork is not new, but it is only recently that entirely portable machines, which are robust enough for archaeological work have become available. It is essential that during fieldwork the means of recording is with the archaeologist, so that details of stratigraphy and artifacts can be recorded in situ. A machine suitable for this purpose must be proof to the elements and rough handling, and should be portable enough to be moved to any part of an excavation. As it is likely to be used in cold weather with gloved hands it should have a sufficiently large keyboard for this to be possible. It must have adequate storage for several hours recording, or be linked to another machine with this facility To guard against loss of data due to machine failure (and to satisfy the needs of paper hungry archaeologists) it must be possible to produce paper output at the time of input.

Software for data entry must be compact, whilst being able to support a complex data structure. For instance a single context may have as many as 100 attributes. Finally it must be possible to interface this small machine to one which will provide facilities for mass storage and manipulation.

Whilst there are now several portable computers with a variety of facilities the only on which appears to be immediately suitable for archaeological purposes is the Husky, but with a basic price of around £2000 this is clearly beyond the means of most archaeologists. Work at West Heslerton is currently asessing the Sharp PC1500 (Powlesland 1983), which has been succesful with racing yachtsmen, and is priced at well under a quarter of the cost of the Husky. At the National Maritime Museum an Epson HX-20 is being tested. Waterproofing is achieved by sealing the Epson in a plastic bag. All functions of the machine can be controlled through the bag; the machines full sized keyboard aiding data entry in the field. Additionally the Epson is likely to prove to be extremely useful in the museum, for stocktaking and other activities where it is not possible to have a full sized VDU.

Fieldwork this summer will test the utility of both these machines, and it is likely that, as has been suggested by Nagle and Wilcox (1982) the small computer will be extremely useful in the museum.

Analysis during and after fieldwork

Archaeology has traditionally proceded by collecting data in the field, and examining it when excavation has finished. However experience suggests that even simple listings of finds and stratigraphy can be extremely useful to the fieldworker (Booth et al, forthcoming). The potential for such analyses to aid in planning the strategy of fieldwork are great. In particular if the best use is to be made of the many techniques for sampling which are available optimal sample size must be determined while work is proceeding (Gregson 1982).

For data collected during fieldwork to be analysed it must first be put into the computer. It may have already been collected by a small computer, in which case adequate interfacing will be necessary between the small machine used in the field, and the larger machine which will perform the analyses. Interfacing is apparently a simple procedure, accomplished with a physical exchange of data (whether by disk, tape, wire or radio), and a program which will make this data intelligible to the machine which is to use it. Unfortunately the two machines concerned may not have been designed with this particular union in mind, and whilst interfacing is rarely impossible it is often difficult. Alternatively the data may be held in paper form, and it must be transcribed. As data entry is likely to be the most costly part of computerisation (Nagle and Wilcox 1982) the identification of the most appropriate means of input is important, and the relative merits of interactive data entry programs versus simple text input will need to be asessed.

Data analysis during fieldwork can broadly be divided into simple listings of catalogues and indexes, and more complex analyses. In addition the ad hoc interrogation of the data to answer specific questions may be necessary. Hardware for all of these must have sufficient storage for a significant portion of the data to be held on line at once, as it is extremely inconvenient to have to work with a large number of floppy disks.

Software must allow efficient data entry, and its flexible expansion. At this stage information will need to be frequently updated, deleted, and interrogated; a direct access file structure would be most appropriate. At present there seems to be little software for complex analyses on microcomputers, so it is likely that the researcher requiring these techniques will need to interface

it to a larger machine offering the necessary packages. The facilities which are at present available on microcomputers include a number of commercially produced and 'home grown' database packages, the GOS programme for catalogues and indexes (Museum Documentation Association 1980), some rudimentary statistics, and the STRATA programme for stratigraphic analysis. The overall impression is that whilst computing hardware has dramatically improved there is a shortage of useful software.

In the Archaeological Research Centre at the National Maritime Museum a database package developed in the museum is run on a Cromemco CS-1H microcomputer. This is a 64K CP/M system, with a 400K floppy disk drive and 5mb hard disk. The program package is based on that developed at Maxey (Booth 1980), and is capable of tagging the data so that it may be processed by GOS.

Requirements for information handling after fieldwork are similar to those whilst it is proceeding. Data will need to be added and edited, and some redundant information may need to be abandoned. Similar types of analyses will be necessary, but unlike fieldwork in a remote situation it should be easier to find the appropriate facilities.

Publication and archiving

There is much debate over what constitutes an archive, and what should be published, and with the arrival of microfiche and 'electronic publishing' there is now discussion as to what constitutes proper publication. The Frere report (Ancient Monuments Board 1975), and more recently a working party of the Council for British Archaeology, and Department of the Environment have attempted to establish guidelines. Whatever is eventually agreed it is clear that the totality of data will need to be properly arranged, with facilities for retrieval, and a portion of this data, with discussion will need to be extracted for publication. It has been argued that there is no need to properly organise the archive, as the excavation or fieldwork will never be reinterpreted in such detail. Whatever the merits of this argument it is often necessary to consult the archive, and for this to be a practical operation it must be properly ordered. In its final form the archive may be held in a computer, or output to microfiche or paper.

The requirement for the archive is that it should be held in a compact, intelligible form, with sufficient indexing and cross referencing to make it useful to inquirers. If it is to be held on a computer the data, plus applications programs must be kept. If they are not to be

kept on the original machine there will be problems of interfacing to the new host, which will have to be capable of holding the data and its attendant software. There will also be the need to make regular copies of data and programs to avoid their loss due to the decay of the magnetic media. Despite all these difficulties there seems a good case for maintaining the archive in machine readable form, particularly as it will have cost so much to get the information into the computer in the first place.

Data will need to be extracted from the database for publication, and merged with discussion and illustrations. As word processing is now commonplace for the production of archaeological reports, it should not be difficult for data to be passed to the word processing package to be merged with other text, and vice versa. Having compiled the text to be published there are several options for output, the simplest of which is a listing on paper. Potentially more useful would be the production of the text in the form of 'camera ready copy' by the use of a high quality printer. Text could be transferred directly to the printers on floppy disk, for computer typesetting, and microfiche can be produced directly from the computer record.

For archiving the hardware requirements are for there to be a machine capable of interfacing with the source of the original data, and there must be facilities for running programs to interrogate the data. There will be no need for editing or expanding the data so a more compact record format may be used. Output must be in a form that is easily understood. Data for publication will have to be in a form that can be accessed by word processing packages, and it will be necessary for there to be interfaces with computer typesetting, or a microfiche maker if these are to be used. A high quality printer will be necessary for producing camera ready copy.

The package used at the National Maritime Museum is able to compress data for archival storage, keeping it in sequential rather than direct access format. Data is passed to the GOS package for the production of high quality catalogues and indexes, which may be output in paper or microfiche form.

Conclusion

It is clear from the above discussion that no single computer, or software package is (at present) able to do all the processes required for archaeological data.

Effective communication between machines, and packages is therefore extremely important. It is interesting to note that whilst the 8 inch floppy disks have a standard format most archaeological users have the 5¼ inch variety. These small floppy disks vary from manufacturer to manufacturer, and only rarely is it possible to to transfer data in this straightforward way between mchines of different makes. Secondly the provision of effective software for the analysis of data is important, now that we have effective database systems for collecting and storing information. Many archaeologists do not have access to mainframe computers, and they will require good software to get the most out of their increasingly powerful microcomputers. Finally those of us involved in archaeological computing might expect more guidance from the archaeological community as a whole, as to what to do with these machines. Sadly many machines will spend the bulk of their time engaged in word processing, as their owners have little idea of their potential computing power, and there is not enough effective software for the tasks which are required.

References

Ancient Monuments Board 1975 — Principles of Publication in Rescue Archaeology, London, Department of Environment.

Benson, D.G. and Jefferies, J.S. 1980 — Microprocessors and archaeological records. In Stewart (1980), p5-12.

Booth, B.K.W. 1980 — A general programme for archaeological database management. In Stewart (1980), p13-19.

Booth, B.K.W. — Museum based archaeology: An approach to data management. INSTITUTE OF ARCHAEOLOGY BULLETIN.

Booth, B.K.W., Brough, R.L. and Pryor, F.M.M. — Flexible site data storage: a microcomputer approach. JOURNAL OF ARCHAEOLOGICAL SCIENCE.

Catton, J.P.S., Jones, M.U. and Moffett, J.C. 1981 — The 1965-1968 Mucking excavation computer database. COMPUTER APPLICATIONS IN ARCHAEOLOGY 1981, p36-43.

Flude, K., George, S. and Roskams, S. 1981 — Uses of an archaeological database - with particular reference to computer graphics and the writing up process, COMPUTER APPLICATIONS IN ARCHAEOLOGY 1981, p51-60.

Gregson, M.S. 1982 — The sampling of linear features: an empirical methodology for the determination of sample size. In Ray, K.W. (ed) Young Archaeologist, p125-140, Cambridge.

Pryor, F.M.M. 1980 — Survey Excavation, RESCUE NEWS 21, p3.

Nagle, C, and Wilcox, U.V. 1982 — Optical mark recognition forms in data entry. Some applications. JOURNAL OF FIELD ARCHAEOLOGY, 9, 4, p538-547.

Stewart, J.D. 1980 — Microcomputers in Archaeology, Museum Documentation Association Occasional Paper 4, Duxford, MDA.